Research Paper | Information Technology | E-ISSN No : 2454-9916 | Volume : 3 | Issue : 5 | May 2017

BUSINESS INTELLIGENCE SOLUTIONS IN HEALTHCARE SECTOR

S. S. Kulkarni [1] | Neena Chaudhari [2] | Jishan Mulla [2] | Abha Kamble [2] | Basit Sofi [2]

[1] Professor, Department of IT, SAE, Pune, India-411048.

ABSTRACT

Business Intelligence applications are trending now-a-days in many different fields. Recent new area of using this application is in healthcare. This paper focuses on making use of Web Mining, Data Analytics and/or finding various useful hidden patterns inside data and representing it using various data representation techniques. Data representation forms are Interactive reports, Reports with charts and various graphs, etc. The reports can also be generated in many formats such as Excel, pdf, csv, etc. Also, using these techniques in a single java application to generate various useful reports and using this application on various personal computers as a distributed application is also mentioned in this paper.

KEYWORDS: BI (Business Intelligence), ETL (Extraction, Transformation, Loading), Data Warehouse, Healthcare.

I. INTRODUCTION

Business Intelligence is discipline made up of various related activities that include gathering, storing, analyzing, sharing, querying and reporting of data to make better decisions. BI is used in several sectors such as finance, food industry, fashion industry, pharmaceutical industry, video games, etc.

The study conducted by John Lewis et al [4] is one of the initial studies that proposed careful and perceptive application of Business Intelligence that can transform data into information and information further into knowledge and help to improve outcomes and operational efficiency. Here, the researchers have discovered the issues in traditional systems and applied Business Intelligence to overcome these issues which resulted in reduced costs and improved clinical care quality in Cancer centres. The historical and real-time datasets are combined to improve the outcomes that improved decision making [4].

As far as our research is concerned, a review conducted by Neena Chaudhari et al [1] shows that in India practical/genuine BI computer software products are rearly seen; and the ones present are to be purchased and are platform specific. Our work is based on implementing a BI product which is platform independent and free for everyone to use. The tools and techniques used while constructing this java application will be discussed in the forthcoming sections [1].

II. BUSINESS INTELLIGENCE IN HEALTHCARE

Healthcare is changing rapidly and so is the industry's need for analytics and business intelligence. The lack of Business Intelligence strategy is considered as one of the nine fatal flaws in Business Operations Improvement (BOI) [2]. The application must be looking forward to look for the following:

- Manage patient's population
- Reduce waste
- Improve quality of care
- Improve clinical decisions
- Optimize financial and operational performance
- Allocate scarce resource

The three key benefits that BI realizes with clinical Data Warehouse are:

- Enabling a more efficient, scalable reporting process
- Ensuring consistent data that everyone can trust
- Enabling meaningful, targeted quality improvement.

III. WEB MINING

Web mining is the application of data mining techniques to discover patterns from the World Wide Web (WWW). As the name proposes, this is information gathered by mining the web. It makes use of automated tools to reveal and extract data from servers and web reports, and it permits organizations to get to both organized and unstructured information from browser activities, server logs, website and link structure, page content and different sources.

We carried out web mining and found the dataset through cms.gov. The Centers for Medicare & Medicaid Services (CMS), previously known as the Health Care Financing Administration (HCFA), is a federal agency within the United States Department of Health and Human Services (HHS) that administers the Medicare program and works in partnership with state governments to administer Medicaid, the State Children's Health Insurance Program (SCHIP), and health insurance portability standards. In addition to these programs, CMS has other responsibilities, including the administrative simplification standards from the Health Insurance Portability and Accountability Act of 1996 (HIPAA), quality standards in long-term care facilities (more commonly referred to as nursing homes) through its survey and certification process, clinical laboratory quality standards under the Clinical Laboratory Improvement Amendments, and oversight of HealthCare.gov [5].

IV. DATABASE SERVER AND DISTRIBUTED APPLICATIONS

Business Intelligence applied to any field needs three basic things to be performed on data i.e. ETL (Extraction, Transformation, Loading). For these processes to be performed we need to dump our huge data somewhere that it could be retrieved, processed and again stored. Hence, we have made use of Oracle Database for this purpose. As I think is a worth mentioning topic that all the tools and applications that will be discussed further are installed on Linux operating system, so all the tools and products are available at zero cost .

Elaborating about Oracle databases we are using Oracle Database Express Edition 11g. Oracle Application Express is a web-based software development environment that runs on an Oracle database. It is fully supported and comes standard (at no additional cost) with all Oracle Database editions and, starting with Oracle 11g, is installed by default as part of the core database install. It's a Object- relational database management system written in Assembly language, C and C++. The Oracle RDBMS stores data logically in the form of tablespaces and physically in the form of data files ("datafiles"). Tablespaces can contain various types of memory segments, such as Data Segments, Index Segments, etc. Segments in turn comprise one or more extents. Extents comprise groups of contiguous data blocks. Data blocks form the basic units of data storage.

New users can be added to the Oracle database using the Sqlplus command line shell provided to user with the installation. As this project contains a Test database, which has various dimension tables in a star schema. The Test user acts as the database server for extracting useful data and pass it to the query and report designer engine. As Oracle XE 11g uses web based approach for communicating with user, it takes port 8080 as default port number. Hence after installing if user goes to port 8080 it will take him/her to the Oracle XE page and this facility provided can be exploited as to construct a distributed application which communicates with the server through this port to different computers to extract data.

The idea behind this application is to implementing it machine independent and free. As it is constructed on Linux it is free and using a server and connecting this application to same server every time it asks for data is making it a platform independent application. Exploiting this facility any user may generate reports as per his/her convenience with the same data from the same server with no costs charged on reporting.

V. DATA WAREHOUSE DESIGNING

A data warehouse (DWH) is a system used to store information for use in data analysis and reporting. Data marts are areas of a data warehouses used to store information needed by a single department or even by an individual user.

The ETL process is used to add "new" data to the OLAP system on a regular basis. ETL is short for Extract, Transform and Load. As the name hints, we'll extract data from one or more operational databases, transform it to fit our ware-

house structure, and load the data into the DWH.

Dimensional modelling, which is part of data warehouse design, results in the creation of the dimensional model. There are two types of tables involved:

- Dimension tables are used to describe the data we want to store. Each dimension table is its own category and can have one or more attributes.
- Fact tables contain the data we want to include in reports, aggregated based on values within the related dimension tables. A fact table has only columns that store values and foreign keys referencing the dimension tables. Combining all the foreign keys forms the primary key of the fact table.

A. STAR SCHEMA

The star schema is the simplest model used in DWH. Because the fact table is in the centre of the schema with dimension tables around it, it looks roughly like a star. This is especially apparent when the fact table is surrounded by five dimension tables.

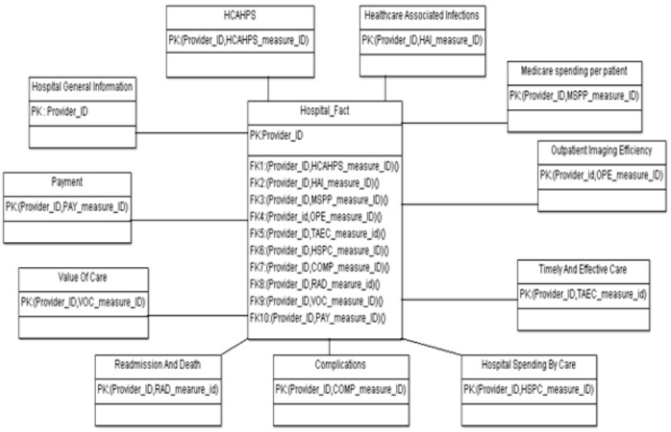


Fig. 5.1. Star schema

After designing our star schema, an ETL process will get the data from operational database(s), transform the data into the proper format for the DWH, and load the data into the warehouse.

The model presented above contains of one fact table and eleven dimension tables. The tables in the model are:

- Hospital_Fact- This table contains references to the dimension tables. Note that all eleven foreign keys together form the primary key of the table.
- Complications- This table contains hospital-level results for surgical complications measures.
- Readmission_and_Death- This table contains hospital-level results for 30-day mortality and readmissions measures.
- Hospital_Spending_by_Claim- This table contains Medicare Spending per Beneficiary breakdowns by claim type
- Hospital_General_Information- This table contains general information on hospitals within the dataset.
- HCAHPS- This table contains hospital-level results for the Hospital Consumer Assessment of Healthcare Providers and Systems
- Timely_and_effective_care- This table contains hospital-level results for Process of Care measures.
- Healthcare_Associated_infection – This table contains hospital-level results for healthcare-associated infections measures
- Medicare_spending_per_patient- This table contains hospital-level Medicare Spending per Beneficiary
- Outpatient_imaging_efficiency- This table contains hospital-level results for measures of the use of medical imaging
- Payment and Value_of_care- This table contains hospital-level results for payment measures and value of care displays associated with 30-day mortality measures.

VI. PENTAHO DATA INTEGRATION

Pentaho Data Integration allows users to intake, blend, cleanse, transform and prepare diverse data from any source. It consists of visual tools to eliminate coding and complexity. Pentaho allows users to combine data from multiple sources into a standard format and make it available for use by different companies.

Pentaho Data Integration (PDI, also called Kettle) is a product of Pentaho which is responsible for the Extract, Transform and Load (ETL) processes. Though ETL tools are frequently used in data warehouses environments, PDI can also be used for many other purposes such as:

- Migration of data between different applications and databases
- Exporting data from databases to flat files
- Loading huge amounts of data into databases
- Data cleansing
- Integrating applications

PDI is very easy to use. Every process is created with a graphical tool where we can specify what to do without writing code to indicate how to do it. Because of this reason, we can say that PDI is metadata oriented.

PDI can be used as a standalone application, or it can be used as part of the larger Pentaho Suite. As an ETL tool, it is the most popular open source tool available. PDI supports a vast variety of input and output formats, which includes text files, data sheets, and commercial and free database engines. Moreover, the transformation capabilities of PDI allow you to manipulate data with very few limitations.

A. EXTRACTION, TRANSFORMATION AND LOADING

In our project, we need to extract data from the database, transform it to standardized format and load it into the data warehouse for generating reports.

1. Extraction:

The source data consists of a large number of tables and these tables consist of a number of columns. Many of these columns do not contribute to the final data that we need on which to perform reporting. Therefore, we need to remove these columns. This can be done by loading the entire table as it is and then removing the required columns. But the better approach is to extract only those columns which contain meaningful data. This is the approach that we have used in our project. By doing this a large amount of time is saved as the tables are huge and by selecting only a select few tables reduces the time to extract the data.

2. Transformation:

The data in our source database consists of data of different data types and different formats. For example, there is numerical data stored in string format. Therefore, we need such transformations in order to get the data into a standardized format. Following are the different types of transformations that we have carried out:

i. Changing of data types:

There are many instances of data type incompatibility in the source data. For example, the numerical data is stored in string format. PDI provides us with options which converts the data from one type into another as per our requirement. There are also instances where the data that we have needs more space, i.e. the data might have 64-bit size and it might need 128 bits. This can also be done here.

ii. Null if transformation:

There are many records in our table whose columns consists of values such as 'Not Available' or 'Not Applicable'. These records are not useful to us and therefore, we convert these values to null so that we can filter them later using the filter transformation. The process of converting values to 'null' is done by the null if transformation. We give a condition such as if value equals 'Not Available' and if the condition matches the records in the table then we change it to null.

iii. Filter transformation:

After converting many records to 'NULL' we get a large number of tables in our database whose columns consist of NULL values and these values are of no use to us. These null values cover a large number of columns, thereby increasing the Cartesian product when we join the tables to form a fact table. Therefore, removal of these null value records is an important step in our project. This task in done by the filter transformation. It filters all records which consist of null values in some specified column and gives us the tables free of null values.

iv. Removal of symbols such as '$',',':

We have many columns in our tables which deal with numerical data in the form of money and these numbers are generally associated with '$' or ',' symbols which requires them to be stored in string format. To process these numbers, we need them to be in integer format and therefore, we first need to remove these symbols.

Here are the examples of the transformations:

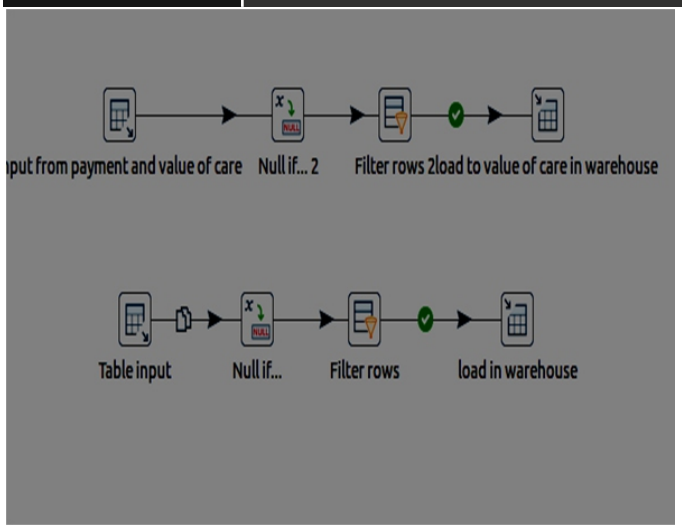


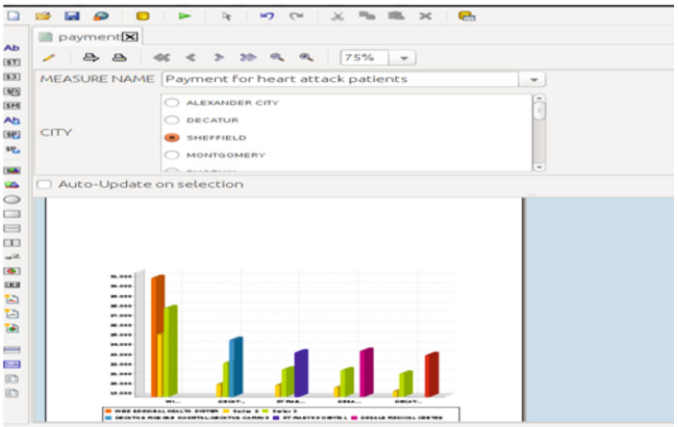
Fig. 6.1. Data Transformation

3. Loading:
The final part is loading the data into the data warehouse. This data is loaded into the Oracle database and consists of the transformed or standardized data. The method we have used is Full Loading which means we are loading all the data into the warehouse at house.

VII. PENTAHO REPORTING

Pentaho Reporting is a suite of tools for creating pixel perfect reports. With Pentaho Reporting you are able to transform data into meaningful information tailored to your audience. You can create HTML, Excel, and PDF, Text or printed reports. If you are a developer, you can also produce CSV and XML reports to feed other systems [6].

A. Reports in Pentaho:
In its most basic form, a report is a document that contains information for the reader. When speaking of computer generated reports, these documents refine data from various sources into a human readable form. Report documents make it easy to distribute specific fact-based information throughout the company.



Fig, 7.1 Generated report for hospital dataset

The report we have generated is report for payment and value care of hospital. We have generated a 3-Dimensional bar chart for this. Similary, many other types of charts such as pie cgart, line chart, waterfall chart, etc. can be generated and the report can be made more interactive. Also, these reports can be saved in different formats such as pdf, text, html, csv, rft and many more. Also we can mail these reports to specified users.

CONCLUSION

Thus we have constructed a mechanism using business intelligence (BI) techniques to extract useful patterns from huge hospital data and represent these pattern data into a interactive report format using various tools and applications. This can help users of this application to very quickly get information regarding their concerns and use this information reach those hospitals and get treated with facilities that they want. Further as we all know that there is no mechanism made to calculate the efficiency of any BI applications. Hence in future our work may have a embedded sentiment analysis module working for receiving users reviews and extract their views and rating for calculating the efficiency of our work.

ACKNOWLEDGMENT

This project is supported by STES' Sinhgad Academy of Engineering. We thank our principal, Prof. K. P. Patil for providing all the resources we are and will be using during the course of this project. We would like to thank our HOD Prof. Abhay Adapanwar for his comments that greatly improved our manuscript.

We thank our guide Mr. S. S. Kulkarni for providing useful insights and expertise that greatly assisted our project. We would also like to thank our external guide Mr. Dinesh Kulkarni for improving an earlier version of the manuscript, although any errors are our own and should not tarnish the reputation of these esteemed persons.

REFERENCES

1. S S Kulkarni, Neena Chaudhari, Jishan Mulla, Abha Kamble, Basit Sofi, "Review on Business Intelligence Solutions in Healthcare Sector," International Journal of Advanced Research in Computer and Communication Engineering Vol. 6, Issue 08, February 2017.
2. Celina M. Olszak and Kornelia Batko, "The Use of Business Intelligence Systems in Healthcare Organizations in Poland," Proceedings of the Federated Conference on Computer Science and Information Systems pp. 969–976.
3. Osama Ali, Pete Crvenkovski and Helen Johnson, "Using a Business Intelligence Data Analytics Solution in Healthcare," Erie St. Clair Local Health Integration Network Chatham, Ontario, Canada K. Elissa.
4. John Lewis, Prof Siaw-Teng Liaw and Pradeep Ray, "Applying Big Data and Business Intelligence Insights to Improving Clinical Care for Cancer," 2015 IEEE International Symposium on Technology in Society (ISTAS).
5. https://www.cms.gov/
6. http://wiki.pentaho.com/display/Reporting/Pentaho+Reporting+-+User+Guide+for+Report+Designer.